

The Importance of Metadata, Conventions and Standards for Climate Data and Long Term Archives

Jessica.Hausman@jpl.nasa.gov

Jet Propulsion Laboratory/California Institute of
Technology

Ted Habermann

HDF Group



Outline

- What is metadata?
- Why care about metadata?
- Conventions and formats
 - ISO 19115, -1, -2, 19157
 - Climate Forecast (CF)
 - Attribute Conventions for Data Discovery (ACDD)
 - Unified Metadata Model (UMM)
 - Open Geospatial Consortium (OGC)
 - SAFE
- Applications of metadata

What is metadata?

- Data that describes the data
- Standard and structured part of scientific documentation
- Provides spatial/projection, temporal, documentation and provenance information
- Collection/dataset – Information of the overall dataset, instrument, platform, algorithms
- Granule – *The smallest aggregation of data independently described, inventoried, and managed.** Swath spatial and temporal bounds, ancillary files
- Variable – min, max, valid, fill values, flags, standard names

Why care about metadata

- Archive
 - Easier to create and validate logs
 - Easier to manage and maintain
- Distribution
 - Interoperability allows access to multiple tools and services to subset, visualize and access data
- User
 - Search spatially, temporally and parameter
 - Improves reuse of data
 - Reformat data
 - Can access the data via multiple services

Citations = lots of metadata

```
@article{WILLIS20131438,  
title = "DORIS and GPS monitoring of the Gavdos calibration site in Crete",  
journal = "Advances in Space Research",  
volume = "51",  
number = "8",  
pages = "1438 - 1447",  
year = "2013",  
note = "Satellite Altimetry Calibration and Deformation Monitoring using GNSS",  
issn = "0273-1177",  
doi = "https://doi.org/10.1016/j.asr.2012.08.006",  
url = "http://www.sciencedirect.com/science/article/pii/S0273117712005467",  
author = "Pascal Willis and Stelios Mertikas and Don F. Argus and Olivier Bock",  
keywords = "DORIS, GPS, Satellite altimetry calibration, Gavdos, Troposphere"  
}
```

Results: 7
(from Web of Science Core Collection)

You searched for: AUTHOR:
(Mertikas) AND TOPIC: (crete)
Timespan: All years. Indexes: SCI-
EXPANDED, SSCI, A&HCI, ESCI.

...Less**Create Alert****Refine Results**

Search within results for...

- Sort by: Date Times Cited Usage Count Relevance More Add to Marked List
- Select Page 5K Save to EndNote online
- 1. First preliminary results for the absolute calibration of the Chinese HY-2 altimetric mission using the CRS1 calibration facilities in West Crete, Greece
By: Mertikas, Stelios P.; Zhou, Xinghua; Qiao, Fangli; et al.
ADVANCES IN SPACE RESEARCH Volume: 57 Issue: 1 Pages: 78-95 Publ
 - 2. DORIS and GPS monitoring of the Gavdos calibration site in Crete
By: Willis, Pascal; Mertikas, Stelios; Argus, Don F.; et al.
ADVANCES IN SPACE RESEARCH Volume: 51 Issue: 8 Pages: 1438-1447
- Full-Text View Abstract
- Full-Text View Abstract

DOI Citation Formatter

Paste your DOI:

10.1016/j.asr.2012.08.006

For example 10.1145/2783446.2783605

Select Formatting Style:

american-meteorological-society

Begin typing (e.g. Chicago or IEEE.) or use the drop down menu.

Select Language and Country:

en-US

Begin typing (e.g. en-GB for English, Great Britain) or use the drop down menu.

Crosscite.org

Format

ISO 19115, -1, -2

- Provides description of the data from the **collection**, granule, variable
- Geospatial information
- Allows you to provide lots of detail on processing, documentation, provenance, ...
 - This also makes it very complex
 - Allows for linkages to documentation
 - Allows for documenting limitations of the data
- Updates every 5 years
- Typically an XML file external to the data file

Climate Forecast (CF)

- Originally created by the modeling community for temporally and spatially gridded data
- Collection (global), **granule** (global) and **variable** (variable) metadata (attributes)
- There is a process to update standard names vocabulary Format updates by request and reviewed by governing body
- NSF project to look at how group hierarchies and swath conventions will be represented in CF (does not mean it becomes part of the convention)

Attribute Conventions for Data Discovery (ACDD)

- Supports discovery of netCDF and HDF data
- Can be used with CF so the data are more interoperable
 - Most of the fields are extra global attributes (collection and granule)
 - Spatial and temporal bounding information
 - Platform/instrument
 - DOI/persistent identifiers

Unified Metadata Model (UMM)

- NASA Earth science metadata model for Common Metadata Repository (CMR)
- All DAACs submit metadata to CMR
- Took into consideration ISO 19115, -1, -2, ECHO10, CF
- Collection, granule, variable and common
- JSON schema or ECHO10 input CMR will produce ISO record
- CMR feeds EarthDataSerach and WorldView

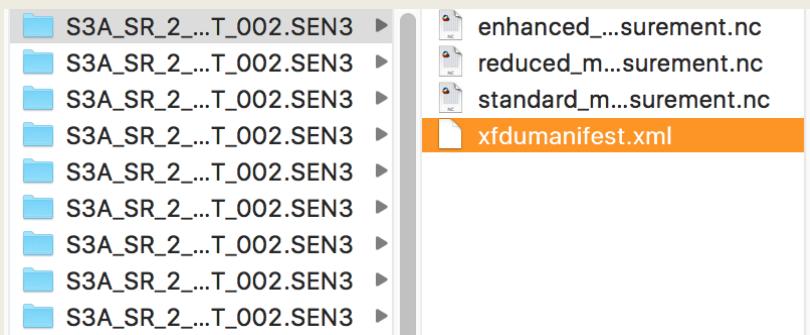
UMM-C	ECHO 10	ISO MENDS (19115-2)	ISO 19115-1	ACDD
DOI/DOI	DOI/DOI	/gmi:MI_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:identifier/gmd:MD_Identifier/gmd:code/gco:CharacterString and /gmi:MI_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:identifier/gmd:MD_Identifier/gmd:codeSpace/gco:CharacterString = gov.nasa.esdis.umm.doi and /gmi:MI_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:identifier/gmd:MD_Identifier/gmd:description/gco:CharacterString contains DOI	/mdb:MD_Metadata/mdb:identificationInfo/mri:MD_DataIdentification/mri:citation/cit:CI_Citation/cit:identifier/mcc:MD_Identifier/mcc:code/gco:CharacterString and /mdb:MD_Metadata/mdb:identificationInfo/mri:MD_DataIdentification/mri:citation/cit:CI_Citation/cit:identifier/mcc:MD_Identifier/mcc:codeSpace = gov.nasa.esdis.umm.doi and /mdb:MD_Metadata/mdb:identificationInfo/mri:MD_DataIdentification/mri:citation/cit:CI_Citation/cit:identifier/mcc:MD_Identifier/mcc:description/gco:CharacterString contains DOI	id
ScienceKeywords	ScienceKeywords	/gmi:MI_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:descriptiveKeywords/gmd:MD_Keywords/	/mdb:MD_Metadata/mdb:identificationInfo/mri:MD_DataIdentification/mri:descriptiveKeywords/mri:MD_Keywords	keywords
	ScienceKeywords/ScienceKeyword	/gmi:MI_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:descriptiveKeywords/gmd:MD_Keywords/gmd:keyword	/mdb:MD_Metadata/mdb:identificationInfo/mri:MD_DataIdentification/mri:descriptiveKeywords/mri:MD_Keywords/mri:keyword	

Open Geospatial Consortium (OGC)

- Mostly Provides and governs service standards for geospatial data
- Useful for georeferenced, raster and vector data
- Web Mapping Service (WMS)
- Web Mapping Tile Service (WMPS)

Standard Archive Format for Europe (SAFE)

- Container format used by Sentinel series
- Zipped package containing normal, reduced and enhanced files in netCDF and manifest metadata in XML
 - Manifest metadata contains geospatial, provenance and checksum information
- Pros
 - Manifest makes archiving and logging easy
 - Packaging makes migration easy
- Cons
 - Unnecessary repetition of data for the user
 - Does not have unique filenames (changing for Sentinel-6/Jason-CS)



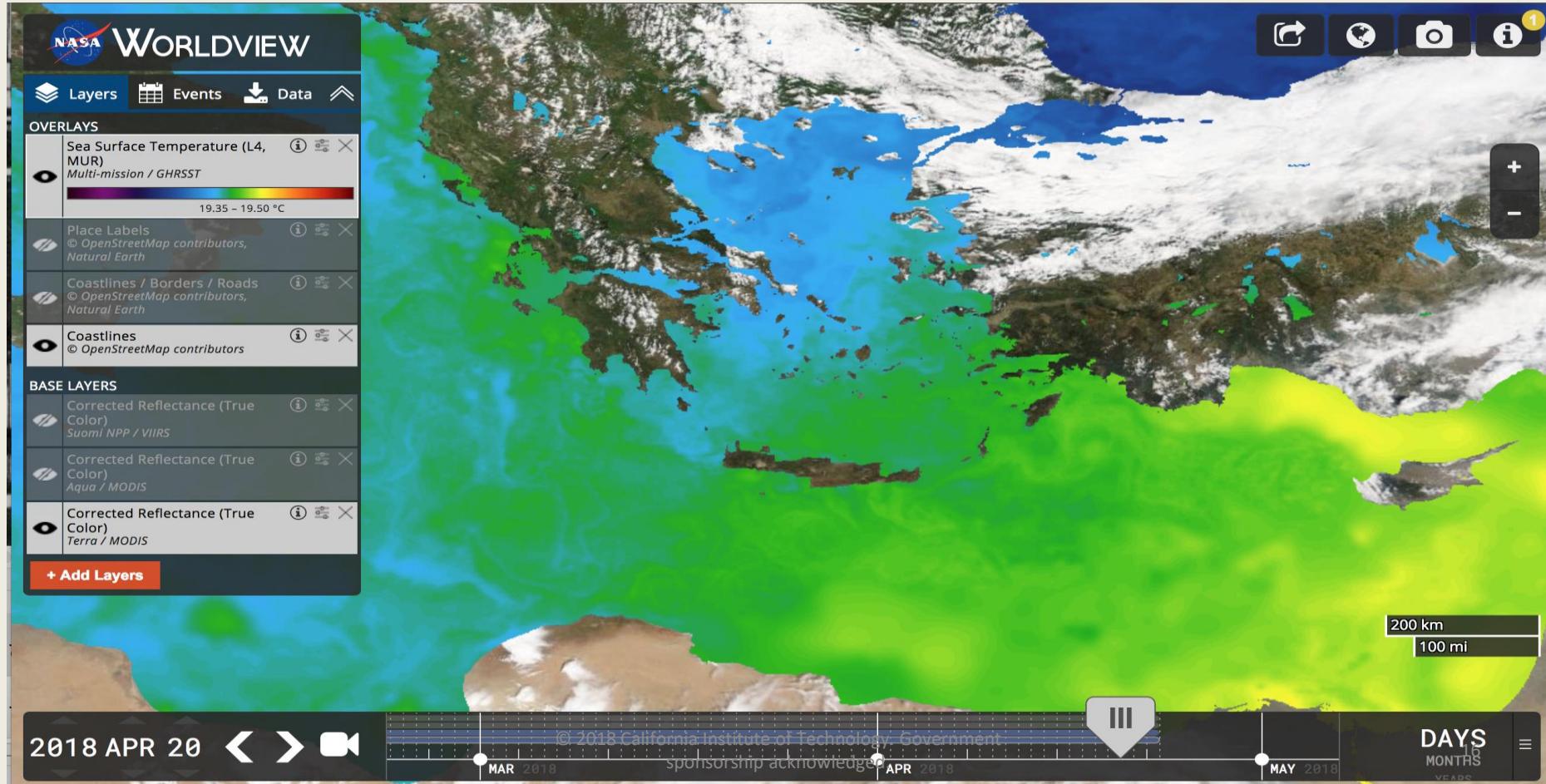
Standard/ Convention	Complexity*	Vocabulary	Concept	Internal or External to data
ISO	Most Complex	multiple	Collection, granule, variable	Internal and external
CF	Somewhat Complex	Standard names	Collection, variable	Internal and external
ACDD	Least Complex	no	Collection, granule	Internal
UMM	Complex	GCMD	Collection, granule, variable	Internal and external

*only in regards to the standards/conventions in this presentation

Applications of metadata

- OPeNDAP – protocol that provides subsetting on time, space and variable (many services use OPeNDAP in the backend for subsetting)
- THREDDS – protocol that subsets and feeds into visualization tools (LAS, Godiva)
- Web services – distributor built protocols or APIs that typically calls OPeNDAP or THREDDS to serve data. Code/script friendly access. Examples/recipes at
<https://podaac.jpl.nasa.gov/forum/viewforum.php?f=85>

PO.DAAC submits MUR UMM metadata -> CMR -> GIBS -> WMTS -> WorldView



Subset and plot data in Matlab via OPeNDAP

```
fnam='http://opendap.jpl.nasa.gov/opendap/OceanTemperature/ghrsst/data/GDS2/L4/GLOB/JPL/MUR/v4.1/2018/108/20180418090000-JPL-L4_GHRSST-SSTfnd-MUR-GLOB-v02.0-fv04.1.nc';
```

```
tic
```

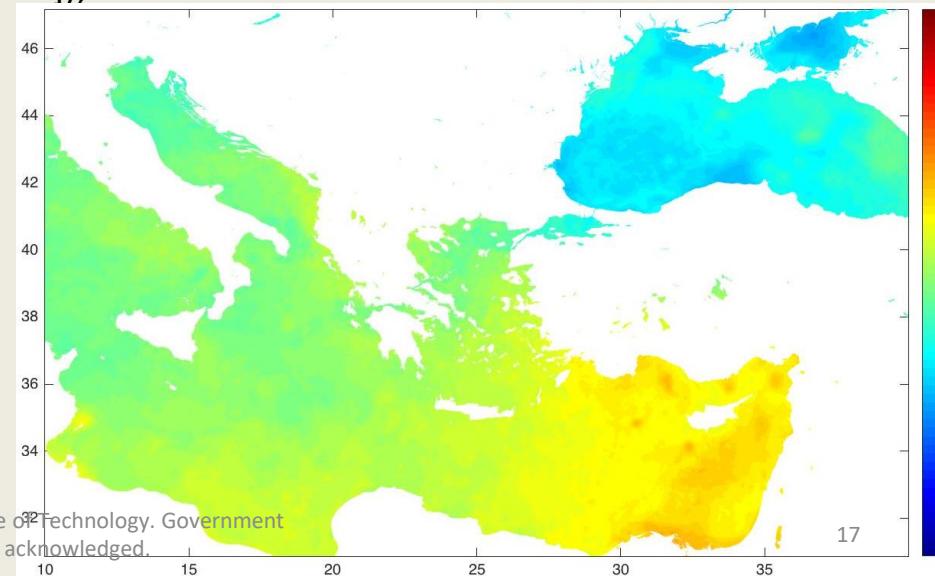
```
lat=ncread(fnam,'lat',12087,1632);  
lon=ncread(fnam,'lon',19000,3000);  
sst=ncread(fnam,'analysed_sst',[19000 12087 1],[3000 1632 1]);  
mask=ncread(fnam,'mask',[19000 12087 1],[3000 1632 1]);
```

```
toc
```

Elapsed time is 87.524977 seconds.

Download via ftp ~7 minutes (over meeting wifi)

File size 384MB



Conclusions

- Metadata can easily link provenance and documentation to the data
- Management and maintenance of the data is easier with metadata
- The knowledge loss when people retire diminishes
- Properly formatted metadata allows the data to be discovered, searchable, subset and reformatted